

## **Drawing amendment**

Applicants have submitted herewith a replacement sheet for the sheet of drawings for this application.

### Explanation of changes

There are no changes to the drawing other than rendering the entire drawing by a draftsman on a computer. The original drawing had a few hand-drawn elements (in particular the set of acoustic models 37 and the horizontal line above it) which are now rendered in a more professional manner. Approval of the replacement sheet of drawing is requested.

## REMARKS

### Specification amendment

The text at page 3 has been amended to change “user” to “system administrator” in order to clarify the distinction between person that administers the system and the actual user (caller). The change does not introduce new matter when the entire passage is read in context.

The text at page 21 has been amended to correct a typographical error (should read “has” not “as”) and change “application” to -- platform -- since the voice command application is mentioned a few words later in the sentence.

### Claim amendment

Applicants have amended claim 1 to improve the form of the claim and to recite that the improvement to the voice command platform relates to instructions (instructions comprising a voice command application) which are stored in a machine readable storage medium. This change makes the claim more clearly recite apparatus subject matter and the previous version of the claim it was not entirely clear the exact class of subject matter of the invention. The instructions include instructions which select a particular acoustic model from a plurality of acoustic model for use by the application. This aspect remains from the previous version of the claim and, as will be described below, is the distinguishing aspect of the invention.

Claim 3 has been amended to improve the form of the claim. The claim consistent with the specification at example 2, page 20.

Claim 8 has been amended to insert a missing word.

Claim 9 has re-drafted claim 4 into independent form for better clarity.

### Anticipation rejection

Claim 1, 3, and 5-7 were rejected as anticipated by Kuroiwa, US Patent 5,960,063.

The Examiner notes that the system of Kuroiwa selects an acoustic model (speech model storage”), but the Examiner errs in considering that such selection is made by a voice command application.

Rather, in Kuroiwa, the selection of one of the acoustic model 7, 8 or 9 is made on a system level, using system level entities, namely the line connection data processor 2 and the switching unit 10. The selection is not made, e.g., by a block of code such as VXML metadata, by a voice command application executing in the system. The reference states as follows:

In response to the line connection data a from the telephone line interface 1 with line connection data acquisition function, the line connection data processor 2 actuates the first switching unit 3 and a second switching unit 6 to select either a first 4 or a second acoustic analyzer 5. Each of the first 4 and the second acoustic analyzer 5 separates the speech data at equal intervals of substantially 10 ms on the basis of a Hamming window of about 25 ms and subjects its data segments to acoustic analysis . . .

It is now noted that the speech may be free from or contain a noise in a particular frequency range of voice signal depending on the route of transmission or the country of the caller, for example, any call from a specific nation in Europe carries such a noise. For handling the former and the latter, the first 4 and the second acoustic analyzer 5 respectively are connected in parallel for selective use.

The first acoustic analyzer 4 analyzes the speech which contains no such noise. The second acoustic analyzer 5 has a notch filter or the like for removing the noise to produce a acoustic vector train from the noise-free speech. The embodiment is not limited to the two acoustic analyzers 4 and 5 shown and three or more acoustic analyzers may be used when three or more noise-imposed speech data are received.

*There are provided three, first, second, and third, reference speech model storages 7 to 9 for saving speech models, e.g. HMMs, defined according to the countries of callers and the routes of transmission. A third switching unit 10 is responsive to a control signal from the line connection data processor 2 for selectively connecting one of the three reference speech model storages 7 to 9 to the speech pattern matcher 11. The speech pattern matcher 11 then compares a speech model d from the selected reference speech model storage with the*

acoustic vector train c transmitted through the second switching unit 6 for speech recognition and delivers its result.

Col 2 line 41 – col. 3 line 10.

The second embodiment of Kuroiwa, Figure 2 is similar to the first:

In particular, the first 12 and the second speech model storage 13 save noise models. A speech may include a silence pause where no sound is made. It is essential for the speech recognition to correctly discriminate the silence pause from the other speech period. However, an intrinsic noise is sometimes imposed on the silence pause when the call travels through a particular route. For identification of such intrinsic noises, their models are saved in the first 12 and the second speech model storage 13. *The switching unit 10 is activated by a control signal from the line connection data processor 2 to selectively connect either the first 12 or the second speech model storage 13 to the speech pattern matcher 11.* The third speech model storage 14 saves speech models for use regardless of the country of a caller and/or the route of transmission. More specifically, the speech models saved in the third speech model storage 14 are identical to those, e.g. HMMs, saved in a speech model storage 34 shown in FIG. 3. The speech pattern matcher 11 identifies the silence pause in a speech from the noise models supplied through the third switching unit 10 and collates the speech with the speech model from the third speech model storage 14 to recognize voice sounds in the speech. A result of the speech recognition is then delivered as an output.

Col. 3 lines 26-50.

Thus, in Kuroiwa, the selection of the acoustic model is not made by instructions in the voice command application itself, as claimed in claim 1, but rather determined on a system level based on analysis of the line connection data and a switching unit. (See also Kuroiwa, Summary at col. 1 lined 56 et seq. and the repeated references to the “telephone speech recognition system”, rather than to a voice command application which is hosted by such a system). Accordingly, the rejection of claim 1 and claims 3 and 5 depending from claim 1 should be removed. The rejection of claims 6 and 7 is moot in view of the cancellation of such claims.

### Anticipation Rejection of Claim 8

Claim 8 is a method claim directed to a method of selecting an acoustic model, including the steps of providing a voice command application with a VXML root document having a block of VXML code and providing in the code a VXML metadata field with an identification of the acoustic model.

Claim 8 is rejected as anticipated by Thomas et al., US patent 7,171,361. The patent does not anticipate because Thomas is selecting a set of *application grammar* to use with the voice command application, *not the acoustic model*. The two are not the same. As stated in numerous places, Thomas mentions that he is selecting the user's preferred idiom "grammar." See col. 5 line 6-22, col. 6, Example 1 (specifying pointers to built in grammar), Example 2 (VXML script explicitly specifies a grammar corresponding to a particular user). The method of Thomas allows the user to specify speech blocks such as dates in the particular idiom (format) the user is used to (e.g., the user's preferred idiom is to say the month, then the day of the month and then the four digit year, instead of a number followed by a month followed by a two digit year, see col. 1 lines 13-17, col. 3 lines 49-57).

The "grammar" that is being referred to in Thomas is a set of words or groups of words which are accepted as valid responses at particular navigation points in the application. Words that are not in this set of grammar are considered "out of grammar" responses, a concept well known in the art. This is what is meant by the term "grammar" in the speech recognition art. See specification, e.g., page 3 line 19 to page 4 line 2. See also the documents attached to this response: Nuance Speech Recognition System Version 7.0 Grammar Developer's Guide, chapter 1 pages 10-11 (pointing out the difference between the

acoustic model and the grammar set in a speech recognition system), Chapter 2, page 13 (giving examples of grammar “Yes’ and “No”); Chapter 3 pages 33-34 (describing choosing an acoustic model set to use with a grammar, clearly indicating that the two concepts are distinct). Speech Recognition Grammar Specification for the W3C Speech Interface Framework, section 1.1 Grammar Processor (“a speech recognizer is a grammar processor with the following inputs and outputs: \* Input: A grammar or multiple grammars as defined by this specification. These grammars inform the recognizer of the words and patterns of words to listen for. . . .”); Technology Reports W3C Speech Recognition Grammar Specification ([The W3C Speech Recognition Grammar Markup Language Specification] “defines the syntax of grammar representation. The grammars are intended for use by speech recognizers and other grammar processors so that developers can specify the words and patterns of words to be listened for by a speech recognizer.”) Wikipedia definition of Acoustic Model: Background (“Speech recognition engines require two types of files to recognize speech. They require an acoustic model, which is created by taking audio recordings of speech and their transcriptions (taken from a speech corpus), and ‘compiling’ them into a statistical representations of the sounds that make up each word (through a process called ‘training’). They also require a language model or grammar file. A language model is a file containing the probabilities of sequences of words. A grammar file is a much smaller file containing sets of predefined combinations of words.”)

Thus, Thomas teaches providing VXML code specifying a grammar, but does specify which acoustic model to use with the speech recognition engine. Accordingly, Thomas does not anticipate claim 8.

Obviousness rejection of claims 2 and 4

Claims 2 and 4 are rejected as obvious over Kuroiwa in view of Thomas. Applicants respectfully traverse the rejection. Claim 4 is cancelled. Claim 2 adds the features wherein the instructions in the voice command application specifying the acoustic model is in the form of a VXML metadata element.

As noted above, Kuroiwa deals with the selection of acoustic model on a system-wide basis though the use of a line data analyzer and a switching element. Kuroiwa does not leave the selection of an acoustic model to the application developer, but rather handles it on a system level. Similarly, Thomas does not even contemplate a voice command application which provides for selection of acoustic models on the application level – e.g., specifying the acoustic model with VXML metadata. Rather, Thomas only teaches selection of the application grammar. Accordingly, neither reference would suggest to one skilled in the art that the voice command application instructions should specify the acoustic model. If anything, such a teaching would conflict with the system-level selection of the acoustic model in the Kuroiwa system. If anything, the two references teach away from claim 2. Accordingly, the invention of claim 2 cannot be obvious over the two references.

Favorable reconsideration of the application is requested.

Respectfully submitted.

McDonnell Boehnen Hulbert & Berghoff LLP

Date: 6/25/07

By: Thomas A. Fairhall  
Thomas A. Fairhall  
Reg. No. 34591

CERTIFICATE OF MAILING

The undersigned hereby certifies that the foregoing Amendment is being deposited as first class mail, postage prepaid, in an envelope addressed to Mail Stop Amendment, Commissioner for Patents, P.O. Box 1450 Alexandria VA 22313-1450 on this 25th day of June, 2007.

Thomas A. Fairhall  
Thomas A. Fairhall